

UNIVERSITY OF TORONTO

ECE1512

DIGITAL IMAGE PROCESSING

Pneumonia Classification Based on Chest X-ray Images

Qing DONG

1005055574

Yiying LAI

1004225607

Due Date:

Dec 2, 2019

Submission Date:

Dec 2, 2019



UNIVERSITY OF
TORONTO

Contents

1	Introduction	2
2	Background	3
2.1	Inception-V3	3
2.2	PSPNet	4
2.3	Transfer Learning	5
3	Methodology	6
3.1	Dataset	6
3.1.1	Chest X-ray	6
3.1.2	MontgomerySet	6
3.2	Problem Setup	7
3.3	Pneumonia Detection	7
3.4	Lung Segmentation	8
3.5	Pneumonia Type Classification	9
4	Results	10
4.1	Model1	10
4.2	Lungseg	10
4.3	Model2	11
4.4	Overall Performance	13
5	Conclusion	15

1 Introduction

Pneumonia is a lung infection that inflames the air sacs. Its symptoms include coughing, fever, chills, sweating, chest pain, nausea or vomiting and difficulty in breathing [1]. While bacteria, virus, and fungi can all cause pneumonia, bacterial and viral pneumonia is contagious. The WHO reports that pneumonia is the single largest infectious cause of child mortality worldwide, accounting for 15 percent of children death under 5 years of age. In 2017, pneumonia killed 808,694 children under the age of 5 [2]. Recently, the risks of pneumonia is increasingly sever in many countries, especially in developing countries. Most of these countries face energy poverty, and as a result, heavily rely on biomass fuels for cooking and heating. This kind of indoors environment increases the chance of contracting pneumonia. In low income regions, the lack of medical resources and the fact that building a health-care system is a long term task that involves educating the public and training the health-care workforce [3] also compound the problem. It is safe to say that pneumonia is an urgent worldwide problem that needs fast and accurate diagnosis.

There are several ways to diagnose pneumonia, including some physical exams such as listening to the patients' lungs with a stethoscope to check if there is any abnormal bubbling or crackling sounds, and blood tests to confirm infection [4]. However, a more accurate and efficient way to diagnose pneumonia is by inspecting chest X-ray images of the patients. Bacterial pneumonia causes the whole lobe to blur, while viral pneumonia causes patterns of blurriness inside the lobe. To save time and cost of diagnosing pneumonia, it would be advantageous to automate the diagnosis. Fortunately, image recognition can serve this purpose. Image recognition is also known as image classification, it takes images as input and outputs a label that corresponds to what is in the images. Deep convolutional neural network models are designed to implement medical image classification. The network has capability to extract more relevant and significant features from the entire image and increase the efficiency of diagnosis. Once the network is trained on a large amount of labeled images, it will be able to detect and classify pneumonia, this process can be treated as pre-diagnose process, followed by confirmation or further investigation of human experts.

In this project, pneumonia classification is decomposed into two stages. In the first stage, the presence of pneumonia is detected, meaning an chest X-ray input image will be classified as normal or pneumonia. In the second stage, the type of pneumonia is determined, meaning images classified as pneumonia in the first stage will be classified as viral or bacterial pneumonia. The technique implemented in this project is deep learning network combined with transfer learning.

This report first describes the background of two deep learning models utilized in this project: Inception-V3 and PSPNet and then briefly introduces transfer learning. Next, it demonstrates the methodology and steps taken for the two-stage problem. Finally it displays the results and analyzes the performance and accuracy of the models, with sample results.

2 Background

Relevant technical background information in this project are explained in this section. A classification neural network InceptionV3 for pneumonia detection of X-ray images, a scene parsing network PSPNet for lung segmentation of X-ray images and a machine learning technique Transfer Learning is discussed here.

2.1 Inception-V3

Inception-v3 was introduced in 2015 to improve computational efficiency and reduce number of parameter in deep convolutional neural networks [5]. It was build using the following four overarching design principles.

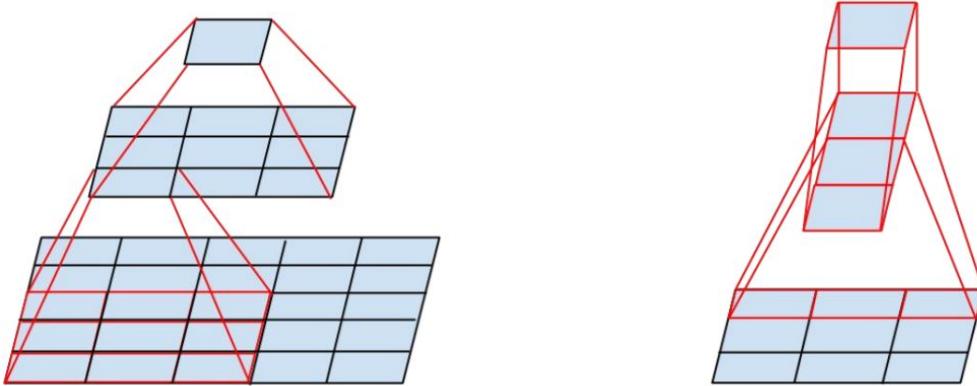
1. Avoid representational bottleneck with extreme compression in representation size (height and width), especially early in the network, meaning it is ideal to gradually reduce the representation size throughout the network.
2. The higher dimensional representations (the more features the convolutional layers extract), the faster the network learns.
3. Spatial aggregation over lower dimensional embeddings because there is a strong correlation between adjacent units.
4. The balance between network depth and width should be achieved.

Techniques used in Inception-V3 include: factorization of convolutions, auxiliary classifiers and efficient grid size reduction.

One type of convolution factorization is to replace a larger convolution with two or more smaller convolution. For example, a 5×5 convolution can be replaced by two consecutive 3×3 convolution, as shown in Figure 1a, enabling the unit at the top to have the same size the field of reception. Another type of convolution factorization is to replace a spatial convolution with two asymmetric convolutions. For example, a 3×3 convolution can be replaced by one 3×1 convolution and one 1×3 convolution, as shown in Figure 1b.

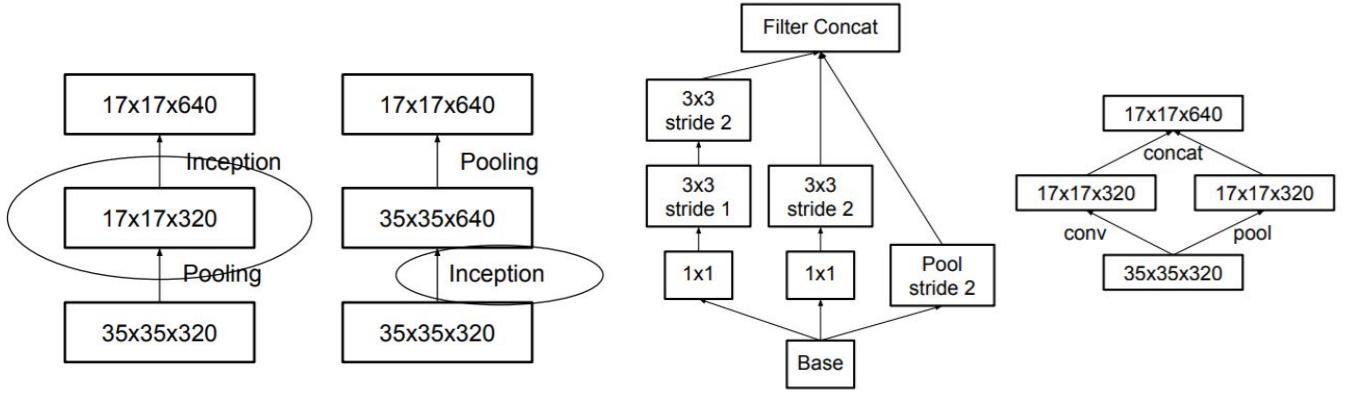
Auxiliary classifiers [6] are also used to combat vanishing gradient problem and act as regularizers. These are classifiers in the intermediate layers, whose prediction are also included in the loss function in the training processing, and they are excluded from the network after training is finished.

Traditionally, grid size reduction is done by performing convolution and pooling in sequence, however, this approach is computationally expensive due to the dominating convolution. If pooling is performed before convolution, the level of computation drops but the first principle of not introducing representational bottleneck is broken. These two non-ideal approaches are shown in Figure 2a. The proposed approach is to perform convolution and pooling in parallel, as shown in Figure 2b.



(a) Replacing a 5×5 convolution with two 3×3 convolutions
(b) Replaced a 3×3 convolution by one 3×1 convolution and one 1×3 convolution

Figure 1: Two examples of convolutions factorization [5]



(a) Non-ideal grid size reduction. The approach to the left breaks bottleneck, while the approach to the right is computationally expensive
(b) Efficient grid size reduction by performing convolution and pooling in parallel

Figure 2: Three examples of grid size reduction [5]

2.2 PSPNet

PSPNet was introduced in 2016 as a scene parsing network [7]. As illustrated in Figure 3, after the features of the input images are extracted by conventional convolutional neural network (a) earlier in the network, the feature map enters the proposed pyramid pooling module. Inside this module, global average pooling is performed on different scales, followed by convolution on each scale, followed by upsampling on each scale. The upsampled features on all scales are then concatenated with feature map (b) and used for pixelwise classification.

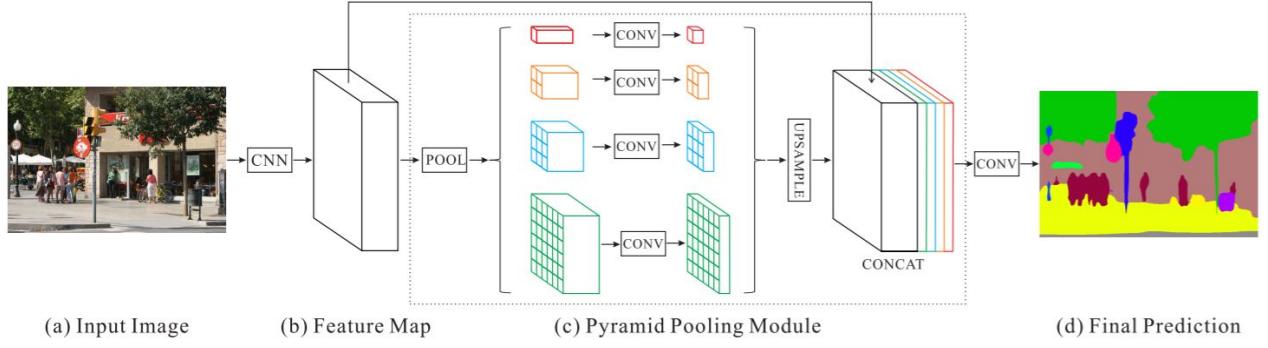


Figure 3: PSPNet architecture [7]

2.3 Transfer Learning

In [11], the definitions of domain and task are given to discuss transfer learning. A domain is represented by $\mathbf{D} = \{\chi, P(\mathbf{X})\}$, where $\mathbf{X} = \{x_1, x_2, \dots, x_n\} \in \chi$ is the training set, χ is the feature space, containing all possible samples, and $P(\mathbf{X})$ is the marginal probability distribution. A task is represented by $\mathbf{T} = \{y, f(x)\}$, where y is the label space, and $f(x)$ a conditional probability function $P(y|x)$.

According to the definition in [11], given a task \mathbf{T}_t based on \mathbf{D}_t we aim to improve the performance of $f_{\mathbf{T}_t}$ by tapping into and transferring the knowledge learned by another task \mathbf{T}_s based on another domain \mathbf{D}_s , where $\mathbf{T}_t \neq \mathbf{T}_s$ and/or $\mathbf{D}_t \neq \mathbf{D}_s$. And often times, the size of \mathbf{D}_t is significantly smaller than \mathbf{D}_s . This is illustrated in Figure 4

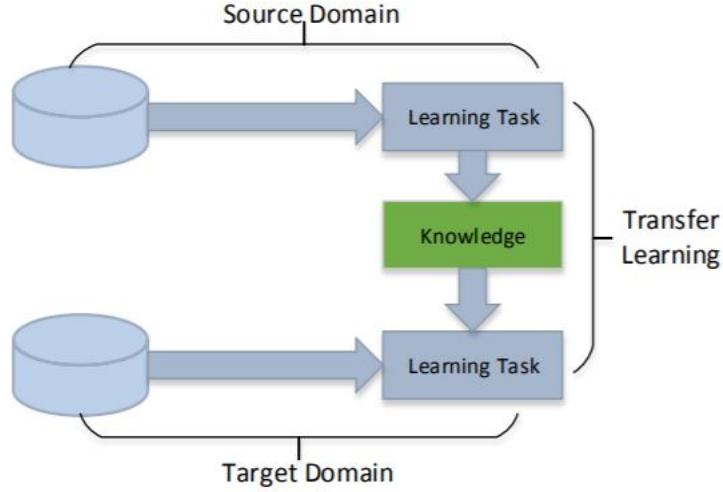


Figure 4: Transfer learning

3 Methodology

3.1 Dataset

3.1.1 Chest X-ray

Two datasets are used in this project. The first dataset is Chest X-ray [8], which contains 5,826 chest X-ray images in total with different size. Three types of X-ray images are present: normal, bacterial and viral pneumonia. Samples X-ray images of normal, viral and bacterial pneumonia type are shown in Figure 5a, Figure 5b and Figure 5c respectively.

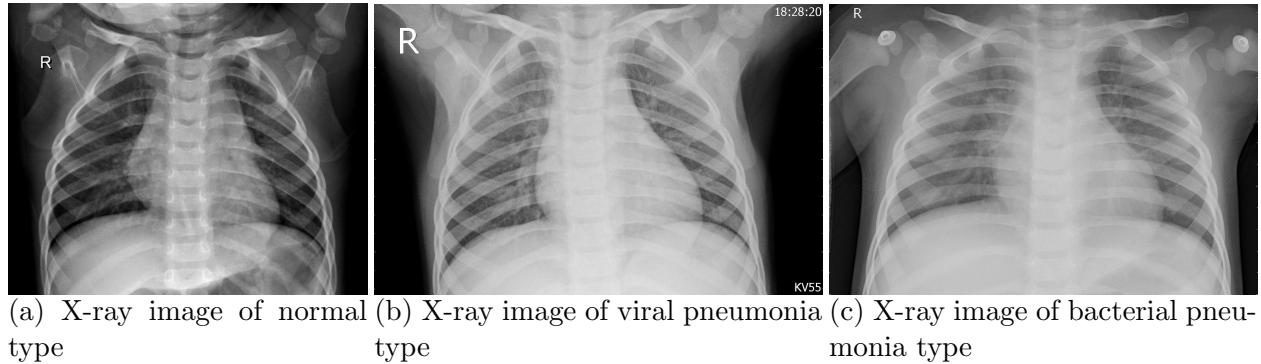
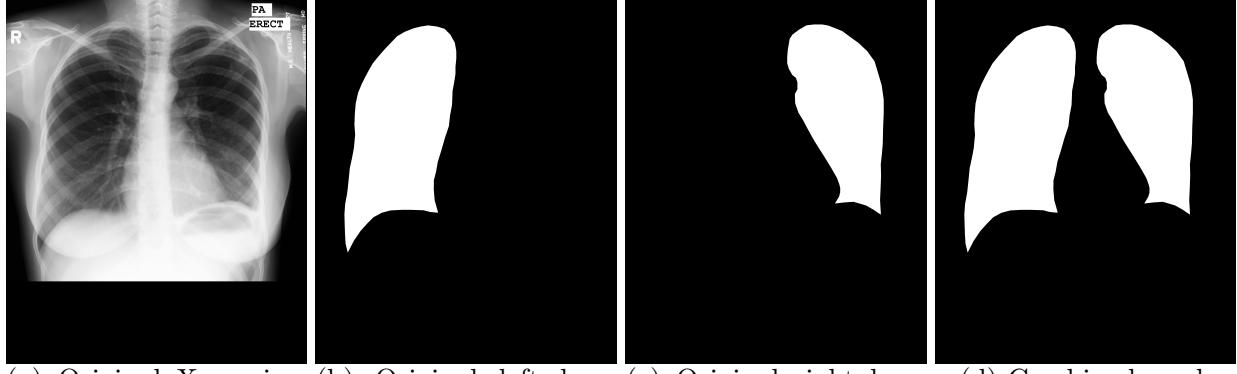


Figure 5: Three sample images from Chest X-ray dataset [8]

The first time the team trained the model on the original dataset, the test accuracy was much higher than validation and training accuracy. It was suspected that the original test set was not a good representation of the dataset, and was too easy to predict, compared with validation and training sets. Therefore the team shuffled all the images in each type and split them into train, validation and test set with a ratio of 0.7: 0.15: 0.15. In stage 1, all of the X-ray images (three types) were used, and only pneumonia X-ray images (viral and bacterial pneumonia) were used in stage 2.

3.1.2 MontgomerySet

The second dataset is MontgomerySet [9][10], which contains 138 X-ray images in total. The folder named CXR.png contains chest X-ray images and the folder named ManualMask contains the corresponding masks of left lungs and right lungs separately in two subfolders. The team combined the left and right lung masks together and save the combined masks in a new folder named Mask. The Mask folder is used for generating labels for the lung segmentation model. Samples of original X-ray images, separate masks and combined masks are shown in Figure 6.



(a) Original X-ray image
(b) Original left lung mask
(c) Original right lung mask
(d) Combined mask

Figure 6: Four sample images from the prepared MontgomerySet [9][10]

3.2 Problem Setup

This project contains two stages. The first stage is to detect pneumonia, where the pneumonia detection model (model1) is trained on the prepared dataset mentioned in 3.1.1.

The second stage is to classify viral and bacterial pneumonia, and only X-ray images classified as pneumonia in stage 1 need to go into this stage. If trained using the original Chest X-ray images, the test accuracy of pneumonia type classification model (model2) was not high enough, thus the team decided to segment lung regions in the original X-ray images and use the segmented X-ray images to train model2, hoping that it will focus more on the region of interest instead of the irrelevant information. The lung segmentation model (lungseg) was trained on MontgomerySet mentioned in 3.1.2, and the improvement of this approach is shown in 4.3.

3.3 Pneumonia Detection

Pneumonia detection model model1 is based on Inception-V3. Thanks to open-source neural network library Keras, the team is able to use Inception-V3 pre-trained on ImageNet [12]. The pre-trained Inception-V3 was originally built to predict 1000 classes in ImageNet, and is not applicable to our task. This is when transfer learning comes into play. Here the task \mathbf{T}_t is to classify X-ray images into normal and pneumonia, for domain \mathbf{D}_t , where the training set is the prepared Chest X-ray dataset. Knowledge learned by task \mathbf{T}_s , to classify images into 1000 classes for domain \mathbf{D}_s , where the training set is ImageNet is transferred to \mathbf{T}_t .

The Inception-V3 in Keras is implemented with an input size of 299×299 , thus the input image target size was set to be 299×299 for the ImageDataGenerator that reads image data and feed to model1.

The pneumonia detection problem is formulated as a two-class categorical classification problem and the last fully connected layer was modified to have two output units, corresponding to two prediction classes normal and pneumonia. The outputs of the last fully connected layer is then fed to a softmax activation layer in the standard way. And the loss function is simply categorical loss of the two output units.

Model1 was compiled with Adam optimizer and trained on the prepared dataset for 17 epochs, which is enough for the training categorical accuracy to reach a plateau. The training images are augmented using horizontal flip, vertical flip and rotation with a range of 10 degrees.

3.4 Lung Segmentation

The lung segmentation model lungseg is based on PSPNet, which is implemented in library keras_segmentation [13], with weights pre-trained on Pascal VOC 2012 dataset. Similarly, the pre-trained PSPNet was loaded and its last fully connected layer was modified to have two output units, corresponding to two prediction classes viral and bacterial pneumonia. The outputs is then fed to a softmax activation layer as well.

The PSPNet in keras_segmentation has an input size of 473×473 , thus the input image target size was set to be 473×473 . Besides, a ground truth segmentation is required by the library to be in a grayscale image, where the value of each pixel is the class index of that pixel. In the case of lung segmentation, a ground truth segmentation is a grayscale image of size 473×473 , where a pixel value of 1 means the pixel should be classified as of lung class, and a pixel value of 0 means the pixel should be classified as of background class. An sample ground truth segmentation image is shown in Figure 7. The segmentation image appears dark because its values are all 0s or 1s, corresponding to the lowest intensity value and second lowest intensity value in a grayscale image with value type unit8.

The segmentation problem is formulated as a two-class categorical classification problem and the last fully connected layer was modified to have $473 \times 473 = 223,729$ output units, corresponding to two prediction classes lung and background. The outputs of the last fully connected layer is then fed to a softmax activation layer and the loss function is simply categorical loss of the 223,729 output units.

Optimizer used is RMSprop. Lungseg was compiled and trained on the prepared dataset for 1 epoch to avoid overfitting.



Figure 7: Ground truth segmentation image of 6a

3.5 Pneumonia Type Classification

Like model1, pneumonia type classification model2 is also based on Inception-V3, and the setup is almost identical, with input size of 299×299 , two output units, two-class categorical loss, Adam optimizer and type types of data augmentation.

After lungseg was trained on MontgomerySet, it is used to produce segmented images for model2.

4 Results

4.1 Model1

As shown in Figure 8, the validation accuracy tracks training accuracy and validation loss tracks training loss. Also, training accuracy increases steadily to above 95% while training loss decreases steadily to below 0.1%, indicating an effective training.

The classification report of model1 is given in Table 1. The high f1-score accuracy can be explained by the fact that X-ray images of both bacterial and viral pneumonia type have opaque area in the lung regions in different degree, and model1 learned to classify normal images from pneumonia images using this distinct characteristic.

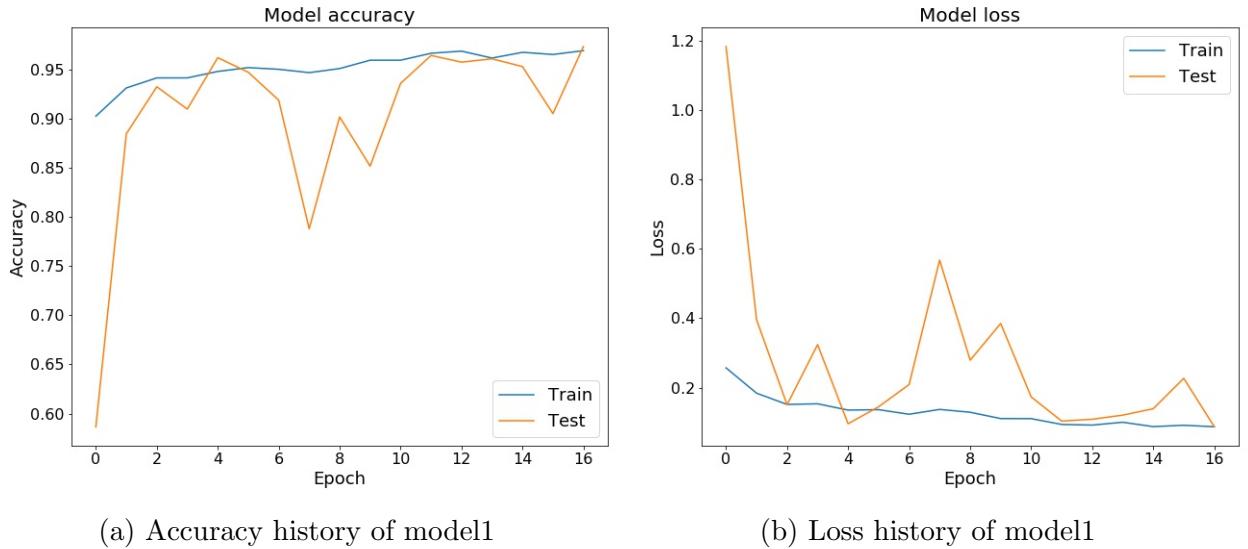


Figure 8: Training history of model1 for 17 epochs

Table 1: Classification report of model1

	precision	recall	f1-score	support
normal	0.94	0.76	0.84	238
pneumonia	0.92	0.98	0.95	641
accuracy			0.92	879
macro avg	0.93	0.87	0.90	879
weighted avg	0.92	0.92	0.92	879

4.2 Lungseg

Samples of lung segmentation on chest X-ray images are shown in Figure 9. Figure 9a is a sample image of normal type, and Figure 9d is its segmented version. The segmentation result is good on this sample because the two lungs are included without missing region.

Figure 9b is a sample image of viral pneumonia type, with an inferior segmentation (Figure 9e) where part of the left lung and right lung is missing. For Figure 9c, its lung segmentation fails miserably, as shown in Figure 9f.

Lungseg underperforms at Figure 9e and Figure 9f because it is trained on Montgomery-Set, X-ray images in which come from patients with tuberculosis and people without tuberculosis nor pneumonia. This means the dataset is biased towards data points that have clear lung regions, so models trained on such dataset give inferior and even bad results for images with opaque lung region.

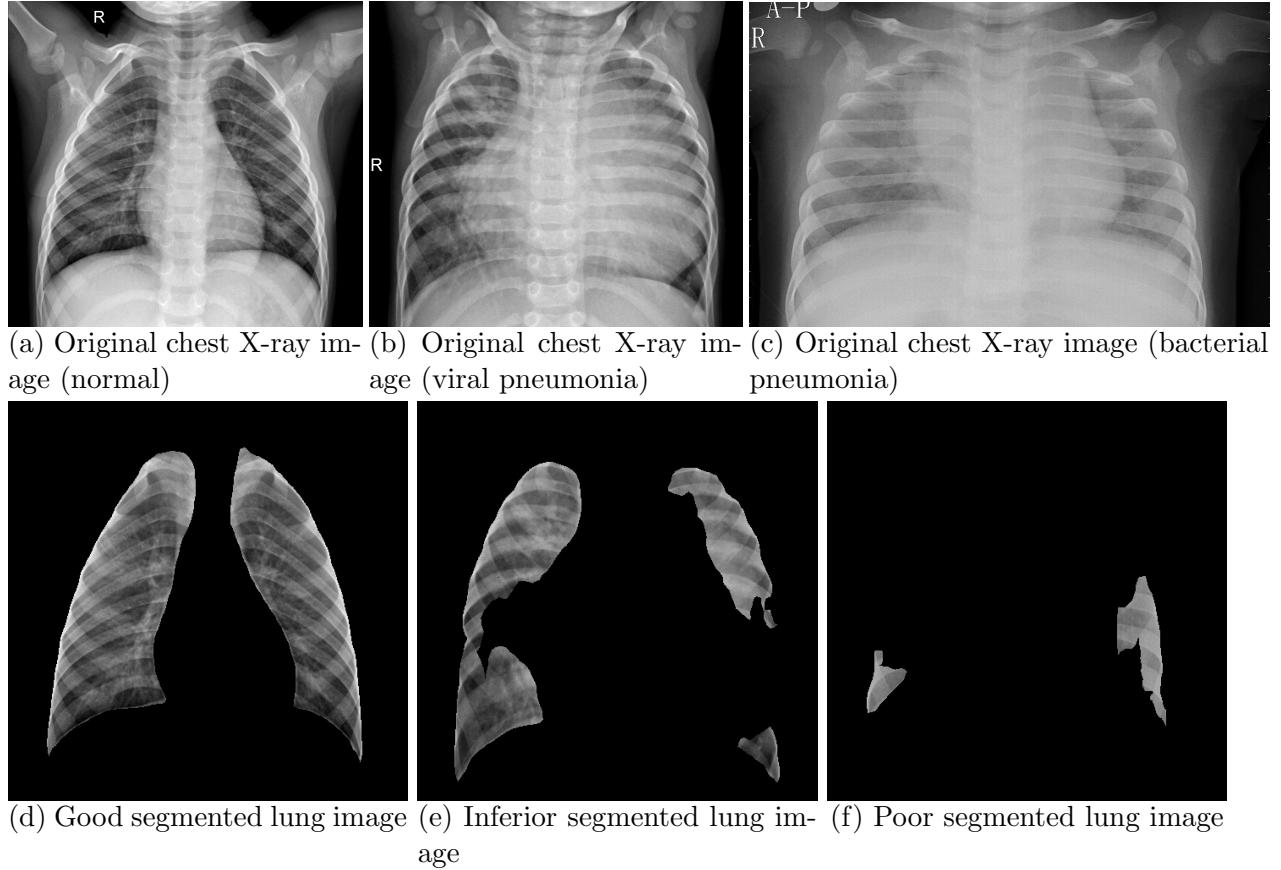


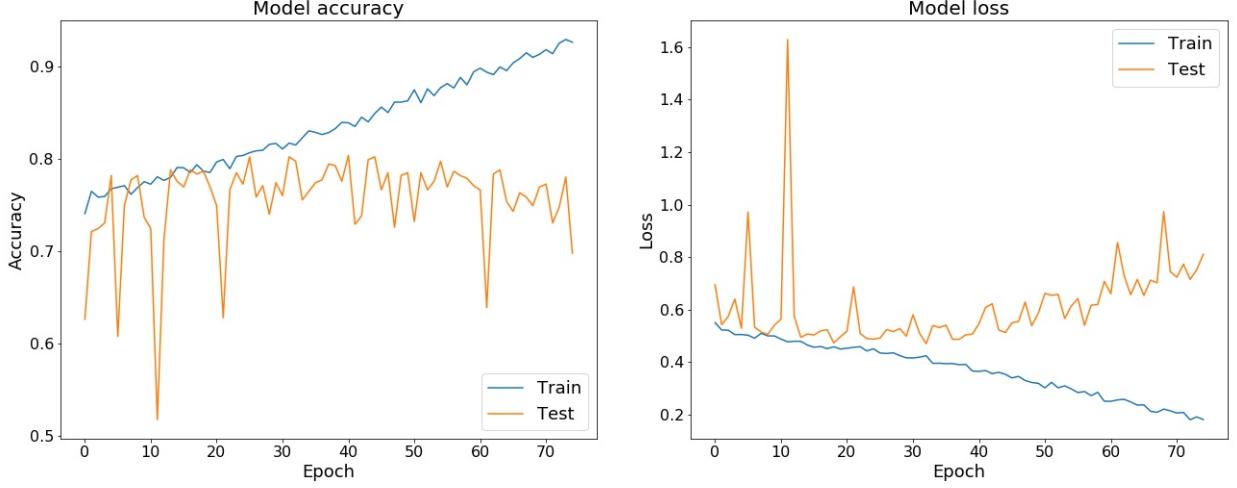
Figure 9: Samples of original X-ray images and segmented images

4.3 Model2

The training history of model2 using original X-ray images is shown in Figure 10. It is clear that training accuracy and validation accuracy diverge after a couple epochs, and the training loss and validation loss diverge after around 20 epochs, both indicating the presence of overfitting. The classification report for model2 trained using original images is shown in Table 2, with a f-1 score accuracy of 70%.

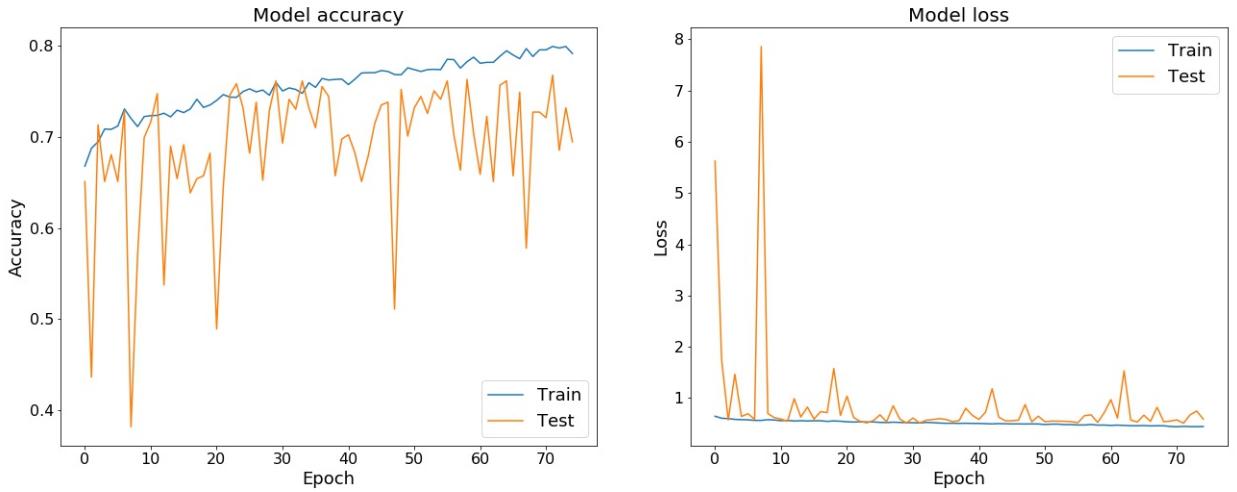
If trained using images segmented by lungseg, model2 is able to avoid overfitting, as shown in Figure 11, where validation accuracy and loss track training accuracy and loss. And the f1-score accuracy on the same testing set is 71%, as shown in Table 3.

Overall, the test accuracy of model2 trained on original dataset is 69% while the test accuracy of model2 trained using segmented images is 72%. A 3% improvement in test accuracy can be found between the two.



(a) Accuracy history of model2 trained with original X-ray images (b) Loss history of model2 trained with original X-ray images

Figure 10: Accuracy and loss history of model2 trained with original X-ray images in 75 epochs



(a) Accuracy history of model2 trained with segmented X-ray images (b) Loss history of model2 trained with segmented X-ray images

Figure 11: Accuracy and loss history of model2 trained with segmented X-ray images in 75 epochs

Table 2: Classification report of model2 trained on unsegmented images

	precision	recall	f1-score	support
bacterial pneumonia	0.82	0.68	0.74	417
viral pneumonia	0.55	0.73	0.63	224
accuracy			0.70	641
macro avg	0.69	0.70	0.69	641
weighted avg	0.73	0.70	0.70	641

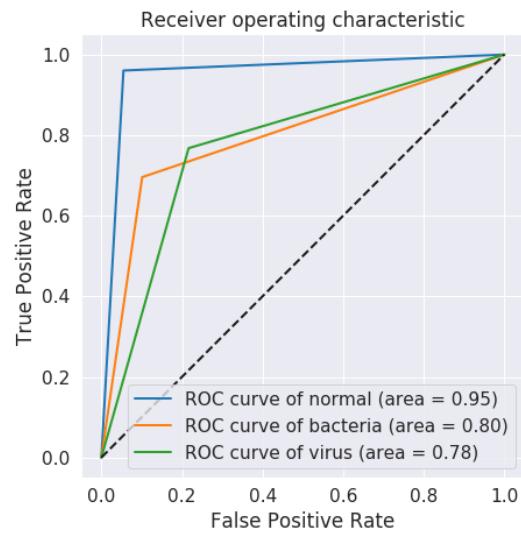
Table 3: Classification report of model2 trained using segmented images

	precision	recall	f1-score	support
bacteria	0.79	0.75	0.77	417
virus	0.57	0.63	0.60	224
accuracy			0.71	641
macro avg	0.68	0.69	0.68	641
weighted avg	0.71	0.71	0.71	641

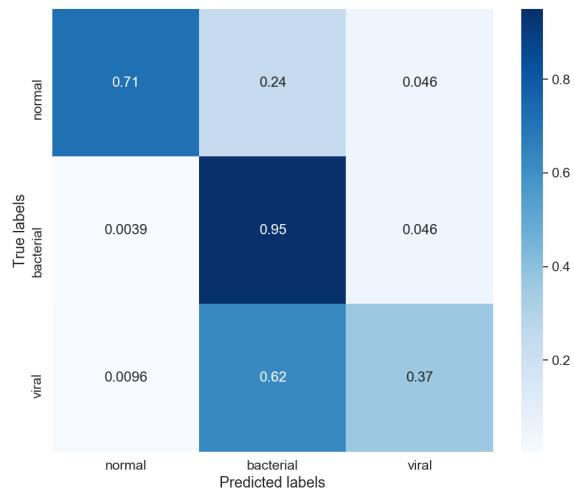
4.4 Overall Performance

The full procedure of classifying a chest X-ray image is as followed: the image to feed the image into model1, if it is classified as of normal type, 'normal' is its final prediction. If it is classified as 'pneumonia' instead, it will be fed into lungseg to produce segmented image. The segmented image will then be fed into model2, where it is further classified as 'bacterial pneumonia' or 'viral pneumonia'.

The overall performance on test set is given in Figure 12. Figure 12a shows its the receiver operating characteristic while Figure 12b shows its confusion matrix.



(a) Receiver operating characteristic of the system



(b) Confusion matrix of the system

Figure 12: Overall performance of the system

5 Conclusion

In this project, three models were created to detect and classify pneumonia. Model1 is trained with normal and pneumonia images from chest X-ray dataset, and its test accuracy is higher than 90%. Lungseg is trained on MontgomerySet to segment lung region in the X-ray image. In general, lungseg performs better on normal chest X-ray images than on pneumonia type chest X-ray images. Finally model2 is trained using both original and segmented images. As shown in 4.3, model trained using segmented images has higher test accuracy than the model trained using original images.

The future work of this project is to increase the accuracy of classifying viral type pneumonia and bacteria type pneumonia by improving the performance of lungseg. As discussed in 4.2, it will benefit from being trained on an unbiased dataset. The proposed solution is to train the lungseg model with some pneumonia chest X-ray images so it would not ignore the opaque lung parts. Also, it is desirable to try other common used models such as VGG16 and Resnet50.

References

- [1] Pneumonia - Symptoms and causes. (2018). Retrieved 1 December 2019, from <https://www.mayoclinic.org/diseases-conditions/pneumonia/symptoms-causes/syc-20354204>
- [2] "Pneumonia", World Health Organization, 2019. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/pneumonia>. [Accessed: October 30, 2019].
- [3] V. Narasimhan, H. Brown, A. Pablos-Mendez et al., "Responding to the global human resources crisis", *the lancet*, vol. 363, no. 9419, pp. 1469-1472, 2004. [Accessed October 30, 2019].
- [4] "Pneumonia - Diagnosis and treatment - Mayo Clinic", mayo clinic, 2019. [Online]. Available: <https://www.mayoclinic.org/diseases-conditions/pneumonia/diagnosis-treatment/drc-20354210>. [Accessed: November 20, 2019].
- [5] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2015). Rethinking the Inception Architecture for Computer Vision. Retrieved 30 November 2019, from <https://arxiv.org/abs/1512.00567>
- [6] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., & Anguelov, D. et al. (2014). Going Deeper with Convolutions. Retrieved 1 December 2019, from <https://arxiv.org/abs/1409.4842>
- [7] Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2016). Pyramid Scene Parsing Network. Retrieved 1 December 2019, from <https://arxiv.org/abs/1612.01105>
- [8] Large Dataset of Labeled Optical Coherence Tomography (OCT) and Chest X-ray Images. (2018). Retrieved 1 December 2019, from <https://data.mendeley.com/datasets/rscbjbr9sj/3>
- [9] Candemir S, Jaeger S, Musco J, Xue Z, Karargyris A, Antani SK, Thoma GR, Palaniappan K. Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration. *IEEE Trans Med Imaging*. 2014 Feb;33(2):577-90. doi: 10.1109/TMI.2013.2290491. PMID: 24239990
- [10] Jaeger S, Karargyris A, Candemir S, Folio L, Siegelman J, Callaghan FM, Xue Z, Palaniappan K, Singh RK, Antani SK. Automatic tuberculosis screening using chest radiographs. *IEEE Trans Med Imaging*. 2014 Feb;33(2):233-45. doi: 10.1109/TMI.2013.2284099. PMID: 24108713
- [11] Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., & Liu, C. (2018). A Survey on Deep Transfer Learning. Retrieved 1 December 2019, from <https://arxiv.org/abs/1808.01974>
- [12] Applications - Keras Documentation. Retrieved 1 December 2019, from <https://keras.io/applications/#inceptionv3>

- [13] Gupta, D. dust: Image Segmentation Keras: Implementation of Segnet, FCN, UNet, PSPNet and other models in Keras. Retrieved 1 December 2019, from <https://github.com/divamgupta/image-segmentation-keras>